Improved Television Systems: NTSC and Beyond

By William F. Schreiber

After a discussion of the limits to received image quality in NTSC and a review of various proposals for improvement, it is concluded that the current system is capable of significant increase in spatial and temporal resolution, and that most of these improvements can be made in a compatible manner. Newly designed systems, for the sake of maximum utilization of channel capacity, should use many of the techniques proposed for improving NTSC, such as high-rate cameras and displays, but should use the component, rather than composite, technique for color multiplexing. A preference is expressed for noncompatible new systems, both for increased design flexibility and on the basis of likely consumer behavior. Some sample systems are described that achieve very high quality in the present 6-MHz channels, full "HDTV" at the CCIR rate of 216 Mbits/sec, or "better-than-35mm" at about 500 Mbits/sec. Possibilities for even higher efficiency using motion compensation are described.

The existing television broadcasting systems — NTSC, PAL, and SECAM — have spawned large and profitable industries; indeed they may fairly be said to have had a profound influence on modern society. Nevertheless, they are far from perfect. Most of the knowledge exists to improve picture and sound quality so as to equal or exceed that of 35mm motion pictures. If we want it enough, we can have higher resolution, better motion rendition, a wider field of view, worldwide compatibility, and freedom from most defects.

Many, if not most, of the issues involved in improving existing systems or introducing entirely new ones, are primarily political or economic and not technological. The main social value of a television system, economics aside, certainly resides in the programming and not the image quality. Interesting as these issues are, they lie outside the subject of this paper. We are concerned here exclusively with means by which sound and picture quality can be improved, and, to a lesser extent, with the means by which such improvements can be brought to the marketplace.

The psychophysical phenomena and signal-processing ideas underlying much of the following discussion were dealt with in an earlier paper and will only be summarized here.[1]

Why Now?

As equipment has improved and professional viewers have become more critical, the shortcomings of NTSC[a] have become more obvious, at least to TV practitioners. What they accepted 30 years ago, in the full flush of the successful design of a color system that did not obsolete existing receivers, is no longer acceptable. There certainly is no grass-roots demand for improvements, and there is little evidence that these perceived shortcomings are significant to the commercial health of TV. Nevertheless, there may well be commercial opportunities in the provision of significantly improved television services.

Rapid progress in semiconductor development, driven primarily by the demands of the computer industry, has led to more powerful chips and lower prices. This presents the possibility of much more "intelligent" receivers and more efficient use of channel capacity. The existence of such capability, and its utilization in related areas, creates a pressure to make applications to TV, as exemplified by digital receivers.

Another factor creating a favorable climate for change in TV is the improved understanding of TV signal processing. Such processing has already been applied to graphic arts, generally at low data rates, but with excellent results. Demonstrations have been made showing good motion rendition with very few frames per second,[2] elimination of interline flicker by up-conversion,[3] and improved separation of luminance and chrominance by means of comb filters.[4]

No doubt the most important element in creating interest in this subject was the demonstration of the Japanese high-definition television system in 1981, a development that took more than ten years.[5] Orchestrated by NHK, with contributions from many Japanese companies, images have been produced that are comparable to 35mm theater quality. The noncompatibility of that system, and the very large required bandwidth, have sparked developments in a number of other countries, mostly directed to improvements that would be easier to implement.[6]

Limits to Quality in Present-Day Television

The image quality actually observed on home receivers is affected by every physical element of the system, as well as by the system design. Since it is spectrum space that is the limited resource, an argument can be made to ensure that *only* the bandwidth limit the quality. As a practical matter, this is not entirely possible. However, the extent to which other, correctable, factors degrade the quality at present is much larger than seems reasonable.

Interlace and Line Structure

The choice of lines/frame and frames/sec for a given bandwidth trades off spatial and temporal resolution. All modern TV systems use 2:1 interlace in order to raise the large-area flicker rate that results from this trade-off. Alternatively, the use of interlace can be considered an attempt to raise the vertical resolution for a given field rate and number of lines/field. The effectiveness of this stratagem is, at best, limited.[7] Interline flicker at the frame rate is produced along with artifacts due to vertical motion. The former becomes disturbing according to the vertical resolution of the camera, which, typically, is

4.5.1

barely half what it might be with 525 lines/frame.[8] Ironically, if the camera resolution were raised to the alias-free Nyquist rate of 262.5 cycles/picture height (cph), the interline flicker would be unbearable.[b] While the artifacts could be avoided by using progressive scanning, that would also permanently enshrine the lower vertical resolution. As discussed below, a solution lies in separating camera and display scan patterns from that of the channel.

Interference Between Luminance and Chrominance: Resolution Compromises

The ingenious band-sharing principle[9] that made NTSC backward compatible does not, in fact, work perfectly with the kind of processing envisaged by the advocates of that system in the 1950s.[c] This is somewhat puzzling, since comb filters were well known for other purposes at that time,[10] and, in fact, had been discussed by Gray in his 1929 patent, the earliest mention of frequency interleaving.[11] The pioneers seemed much more concerned with making the subcarrier less visible on the existing monochrome receivers (that had the full 4.2-MHz bandwidth called for by the existing standard) than on making color receivers work properly.[12]

Whatever the history, we are all now very familiar with cross color, the appearance of spurious color in regions of high luminance detail where luminance information is interpreted as color, and cross luminance, the moving cross-hatch pattern that develops at the boundaries of brightly colored objects where color information is interpreted as luminance. Cross color cannot be removed easily at the receiver, but cross luminance can be controlled by restricting the luminance bandwidth to well under 3.5 MHz, and often as low as 2.5 MHz. The latter procedure, of course, violates the principle of full backward compatibility.

Even without these defects, which, at least in principle, can be removed by suitable multidimensional filtering at both transmitter and receiver, the system design itself restricts horizontal color resolution to a very low value, typically about one-seventh that of luminance, and very much lower than the vertical color resolution. Although there is no need to have vertical and horizontal resolution identical, a 7:1 disparity is very visible.[13]

Performance of Cameras and Displays

Ideally, the response of cameras and display should be 100% throughout the Nyquist baseband. Because of some peculiarities of camera operation, the vertical response is much poorer than that. The temporal response is determined by camera integration. This blurs moving objects, but does not limit the bandwidth enough to avoid temporal aliasing. Use of a shutter to keep moving objects in focus makes the aliasing worse. What is needed is a camera with high and fully controllable resolution in all three dimensions, a result that can only be obtained by operating a camera of inherently high resolution at a line and frame rate well above that of the channel.

Displays have somewhat different problems from those of cameras. The line structure, which tends to mask fine detail, cannot readily be removed without reducing sharpness. The fact that this seems not to be a serious problem in contemporary HDTV systems is further evidence that the camera and the display, and not primarily the system parameters, limit the vertical response. A corresponding effect is present along the time axis. Phosphor persistence short enough to avoid temporal blurring produces a large fluctuating component of brightness in addition to the (time-average) desired value. With normal frame rates, filtering in the human visual system sufficient to avoid flicker must necessarily reduce the response in the temporal Nyquist band.

These phenomena point up the fact that the camera and display not only limit the system resolution, they must therefore also limit the efficiency of utilization of the channel capacity. As long as this is the case, there is no point in providing channel capacity very much in excess of the capabilities of the terminal transducers.

Transmission Problems

The striking difference in image quality between TV receivers in Japanese and American hotels, both using NTSC, is clear indication that it is not just the system, or even the receivers, that limit image quality. Good pictures require good signals at the receiver terminals. Furthermore, it appears that the higher the resolution of the TV system, the more easily quality is lost by such factors as interference, channel noise, and multipath

transmission. Especially in analog systems, there is no simple remedy to this problem, except to advocate a much higher standard of operation and maintenance of every part of the TV plant, from studio to the home. Should digital transmission ever really become the norm in delivery of signals to the home, the job would be immeasurably simplified. Perhaps that is one good reason to go digital. Important as this problem is, it is not the subject of this article.

Improved NTSC

We firmly believe that the full capabilities of the NTSC system have not yet been exploited. While it might not be capable of "theater quality," if the defects are removed and certain other changes made in the studio and in the receiver, then pictures much better than we have ever seen with this system should be produced.

There are many advantages to improving TV on the basis of the present systems. This evolutionary approach, in which existing receivers may continue to be used, albeit without all of the benefits of whatever improvements are made at the transmitter, ensures the presence of the audience, an element necessary for manufacturers, producers, and broadcasters alike to be willing to make the required investments. It also avoids the politically untenable situation in which viewers find their receivers no longer of use because the broadcasting standards have been changed. The disadvantage of this approach is that there is an upper limit to the quality that can be attained, and this limit may well be below what is desired. In addition, even if full theater quality could be achieved in a compatible manner, it is not clear that a sufficiently large portion of the audience would buy new and more expensive receivers just for the sake of higher technical quality. If this were the case, the economies of scale required to produce sophisticated receivers at acceptable prices might never be reached.

It is possible to envisage a scenario in which compatible improvements were introduced over a period of years, during which time the shape of a final, noncompatible system might emerge. By this time the audience would have become more knowledgeable about quality issues and more discerning in its choices. A line of multistandard receivers could then be developed, capable of receiving new
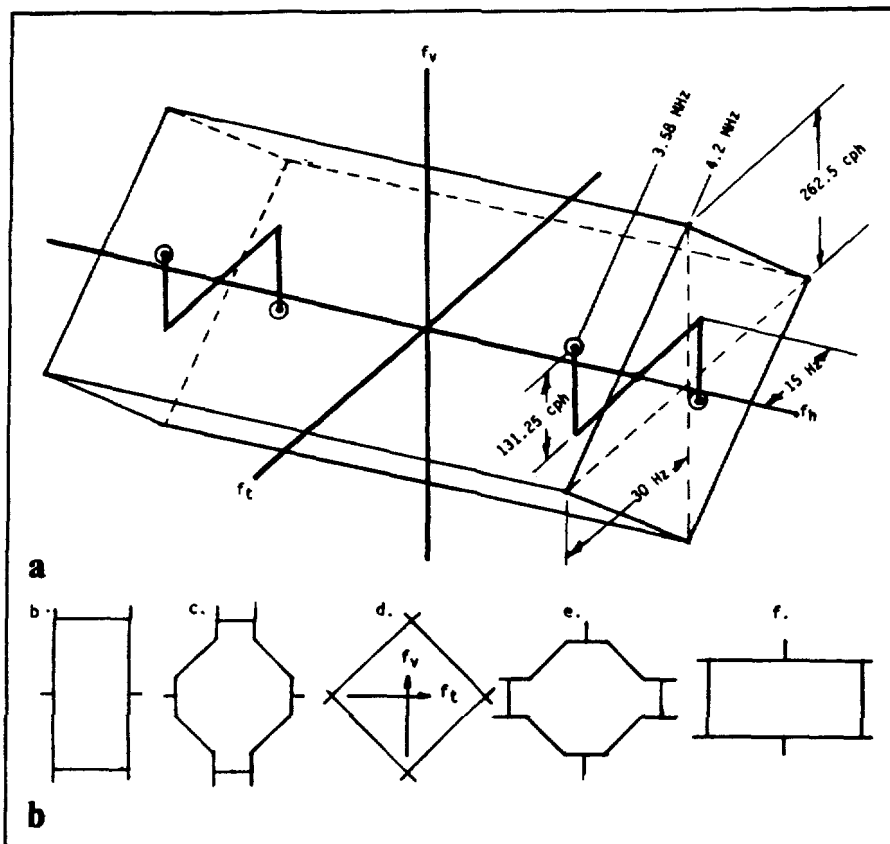
Figure 1. Location of the subcarrier in the 3-D frequency spectrum. The equivalent spatiotemporal subcarrier is located in four of the eight octants. The cylinder with diamond-shaped cross section is the largest symmetrical alias-free baseband possible for luminance with interlaced scanning. Other shapes of the cross section are possible as shown in Figs. 1b to 1f. The shape in Fig. 1b corresponds roughly to full-frame integration, while 1f corresponds roughly to single-field integration, as is more common in current TV cameras. The shape of Fig. 1a is believed to represent a better trade-off of vertical and temporal resolution.
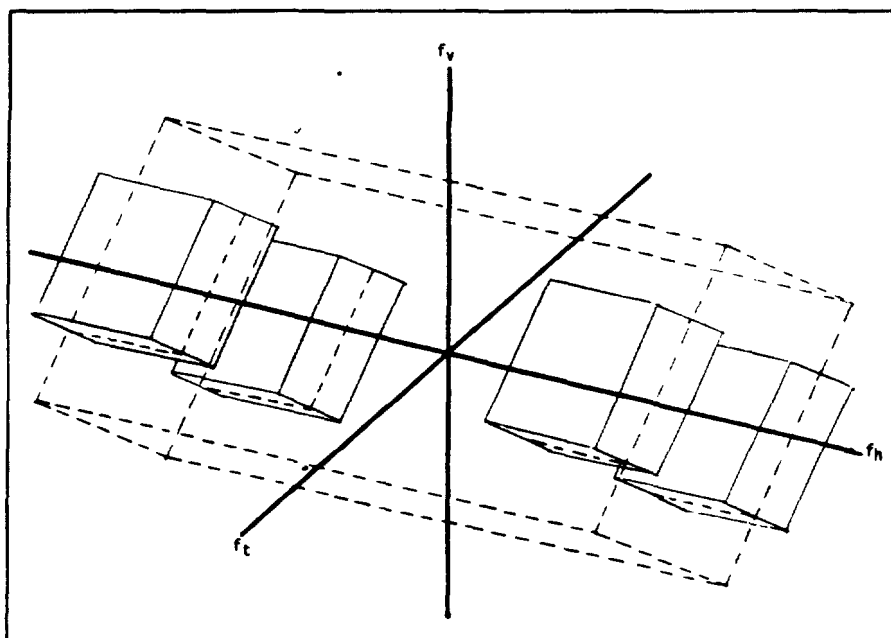

Figure 2. Overlap of chrominance and luminance. The chrominance spectrum is centered around the subcarrier in the same way the luminance spectrum is centered at the origin. The dotted lines show the largest possible alias-free luminance baseband while the four smaller diamond cross-section cylinders show the chrominance with one-half the vertical-temporal bandwidth. Obviously luminance must be made smaller to avoid overlap. Note that the horizontal extent of each component is limited by temporal filtering before and after combining them to form the composite signal. Conventionally, no specific vertical-temporal filtering is done at all.

broadcasts as well as NTSC, and a new broadcasting service with more attractive programs inaugurated. Old receivers would continue to be serviced by down-conversion to NTSC of at least some of the new broadcasts. This would provide an audience for the new programming as well as programs for the existing following. At some point, announced several years in advance, the down-conversion would stop. If it were desired to keep providing some service for the older receivers, special broadcasts of a limited nature could be provided for an additional transition period, as was done in Britain when the shift was made from 405 lines to the CCIR standard and in France when the 819-line system was phased out.

Removal of Cross Effects[d]

Even though Fourier analysis dates from 1822[14] and was ably used by Mertz and Gray[15] in their 1934 analysis of the spectrum of monochrome video signals, it was not properly applied to NTSC signals until 1967.[16] It was not therefore available to the NTSC in 1953, where it would have made much clearer what was involved in proper separation of chrominance and luminance. The full derivation of the spectrum of composite signals is given in Ref. 17 and outlined in the Appendix. At this point we simply note that since the light intensity on the focal plane of the camera is a function of x, y, and t, its spectrum is three-dimensional, the axes being labelled horizontal, vertical, and temporal frequencies. As shown in Fig. 1, the subcarrier corresponds to spatiotemporal harmonics located in four of the eight corners of 3-D frequency space, the chrominance sidebands being located around the subcarrier in the same way that the luminance sidebands are located around the origin.

To be separable, chrominance and luminance harmonics must occupy distinct portions of 3-D frequency space. Normally, the respective video signals are bandlimited horizontally only, and not vertically or temporally except by natural characteristics of the camera. It can readily be seen from Fig. 2 that, if both signals have full bandwidth in these two directions, overlap must be present and cannot be removed by any operation at the decoder. To avoid this irreversible mixing, filtering is required at the transmitter, before encoding.

The simplest filters to use are one-

dimensional, in which case they may be implemented as ordinary lumped-constant networks and applied to the video signals. If luminance is limited to 3 MHz and both chrominance components to 0.6 MHz, i.e., if we give up temporal band sharing, then cross effects are removed. This approach gives up about 30% of the luminance bandwidth and about 50% of the I-signal bandwidth.

Better performance can be achieved with two-dimensional processing, normally implemented with transversal filters[c] using a number of line- and sample-delay elements. As shown in Fig. 3, these can be used to cut the corners out of the 2-D spectrum. Note that it is necessary to limit the vertical resolution of chrominance to perhaps half that of luminance in order to have reasonable vertical luminance resolution at high horizontal frequencies. Note also that "diagonal" luminance resolution must be given up to make room for chrominance. This is generally a good deal less damaging to image quality that giving up horizontal or vertical resolution. Excellent results with such filters have been achieved at INRS-Telecommunications in Montreal.[18] Another approach, which is intended to permit the use of simpler 2-D filters, is to make them adaptive.[19]

With 3-D filters, which require frame delays for implementation, it is, in principle, possible to maintain the full NTSC luminance and chrominance spatial resolution at low temporal frequencies. As shown in Fig. 4, this is done by limiting the spatial resolution of both components at high temporal frequencies. This is generally harmless except in the case of visual tracking of moving objects, in which case high temporal frequency response is required in order to preserve their sharpness. So far, this procedure has never been optimized, so that it is not possible to say how well it may work.

One possibility, with either 2-D or 3-D filtering, is to utilize horizontal resolution of the I component higher than that of NTSC standard by giving up some horizontal luminance resolution at high vertical (or temporal) frequencies. This would cause worse cross effects on existing receivers, which might, in turn, be ameliorated with some kind of adapter.

In designing a new composite system in which 2-D or 3-D filtering were the intended mode of operation, it
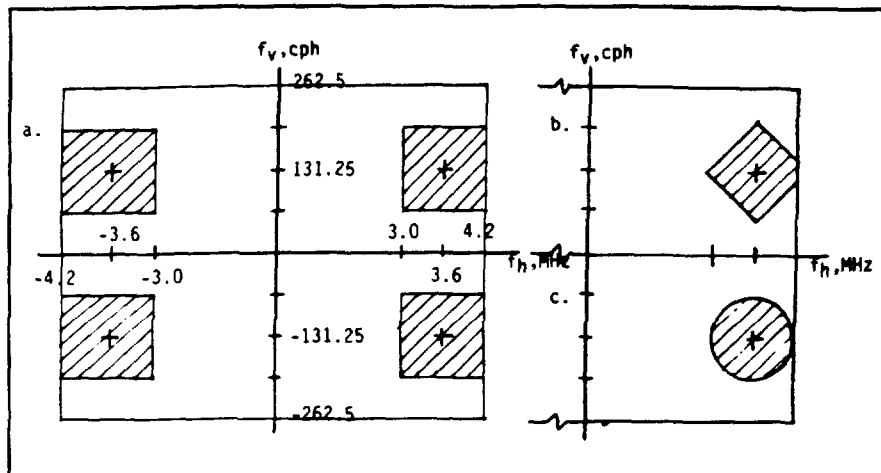

Figure 3. Separation with 2-D filtering. Ignoring temporal effects, the spectrum can be divided in the vertical-horizontal frequency plane as shown, totally eliminating cross effects. Other shapes can be used for chrominance, as shown in Figs. 3b and 3c. For the sake of simplifying the filters, some of the potentially usable spectrum space can be discarded. In Fig. 3a, the vertical chrominance resolution is only 65 cycles/picture height, or one-fourth that of luminance at low horizontal frequency. The vertical luminance resolution at high horizontal frequencies is also quite low.
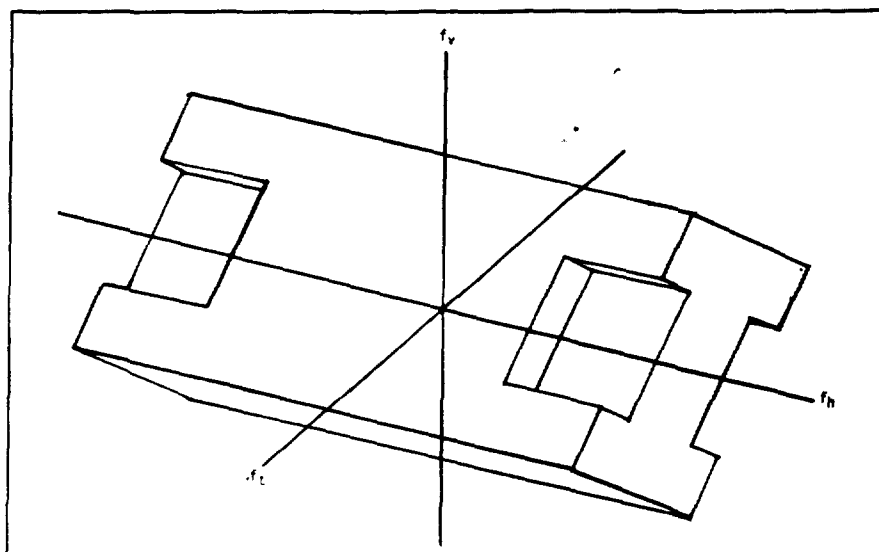

Figure 4. Separation with 3-D filtering. The luminance baseband is shown notched out to make room for chrominance. Many different allocations are possible. In this one, full spatial resolution is achieved for luminance and one-half spatial resolution for chrominance at zero temporal frequency, i.e., for stationary images, while both are reduced at higher temporal frequencies.

might well be possible to use a lower subcarrier frequency, at about 3.2 MHz, to allow double-sideband quadrature modulation of two color components, each of 1-MHz bandwidth. If the vertical color resolution were then limited to half that of luminance, a more symmetrical arrangement would be achieved.

The "Fukinuki Hole"

In the 3-D approach to the spectrum, it is evident that, since only four corners are occupied by chrominance, the other four can perhaps be used for another signal, leaving the luminance spectrum perfectly symmetrical. This requires a second subcarrier, f_h higher

or lower than f_{sc}. Fukinuki[20] first proposed using this extra channel for higher luminance components, and has recently made a demonstration of the process.[21] It should be noted that for the "hole" to exist, the vertical chrominance resolution must be no more than about 131 cph, which is only *half* the vertical Nyquist bandwidth, as shown in Fig. 5. It may well be that the overall image quality would be improved if this spectral space were used, as Fukinuki has proposed, to increase the horizontal luminance resolution rather than to maintain the theoretical vertical chrominance resolution. The latter is low in present-day cameras and, in
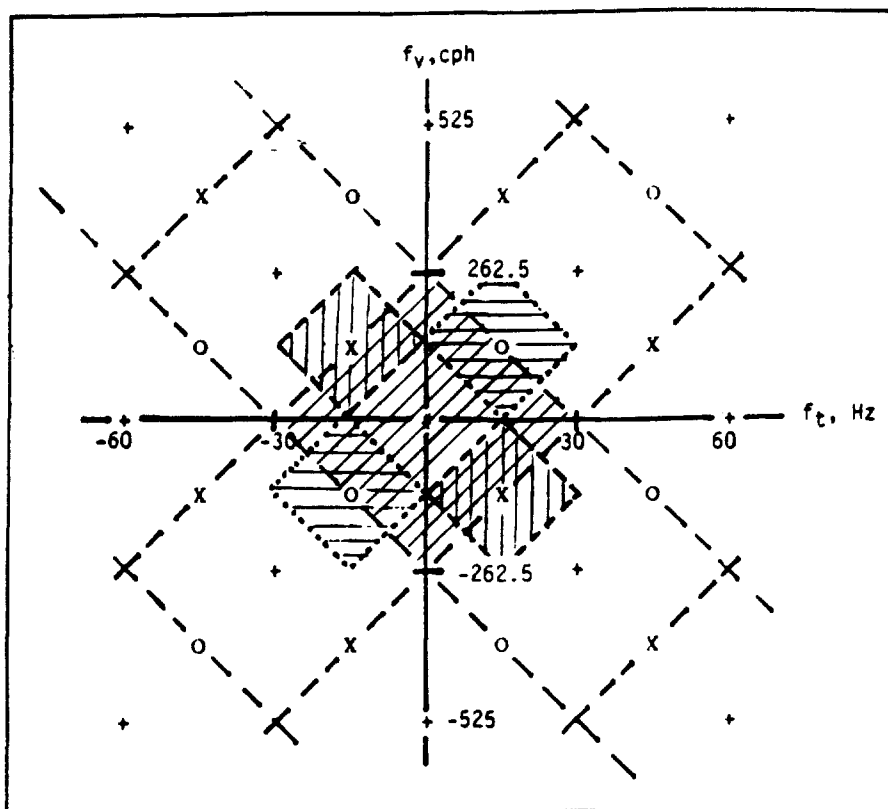
Figure 5. The Fukinuki Hole. Here the spectra are shown in the vertical-temporal frequency plane. The baseband luminance, diagonally cross-hatched, is centered at the origin (#), with replicas centered at the harmonics of the line-scanning frequencies (+). The diamond cross section shown, which is believed to be optimum, provides maximum spatial resolution for stationary images. The color subcarrier and its harmonics are shown by the Xs, vertically cross-hatched, with the chrominance spectrum extent chosen to be one-half of that of the luminance. With this choice, an equal space remains for the Fukinuki subcarrier at the Os, surrounded by the spectrum of an auxiliary signal, horizontally cross-hatched.

any event, cannot be used effectively because of the very low horizontal chrominance resolution. Increased cross luminance and/or cross color is to be expected with this technique on conventional receivers that use notch-filter decoders. Fukinuki, however, reports that this is a small effect.

Separation of Camera and Display Parameters from Those of the Channel

There are a number of important objectives that can be realized more readily, both in improving NTSC and in developing entirely new systems, by separating the scanning standards of the camera and display from those of the channel.[22] One is the elimination of the defects related to interlace. Another is the improvement of the spatiotemporal resolution of the entire system. Both require explicitly controlling the frequency response of the transducers rather than relying on parameters that are an accident of their natural characteristics and of the channel scanning standards. Such ex-

plicit control requires that this response be defined in terms of a group of coefficients — certainly not less than two or three along each axis. At a bare minimum, the sampling density of the transducers must therefore be at least double that of the channel, per axis, or eight times the overall data rate. It is not clear that even this is enough if any reasonable precision is to be obtained, but at least it is a start.

Kell Factor

In analog television systems, it has long been a problem to equate the effective vertical and horizontal resolution, one direction being sampled and the other being rendered in a continuous manner. In the horizontal direction, it has been customary to assume two resolvable elements per cycle of bandwidth, in accordance with the sampling theorem. Equivalent resolution in the vertical direction is found by multiplying the number of active lines by a factor less than one, justified on a number of shaky theoretical and experimental grounds. A

recent study has shown that the Kell factor depends on the pre- and post-sampling filters and can be very close to unity.[23] One of the advantages obtainable with high-rate cameras and displays is the achievement of a high Kell factor.

Temporal Performance

The role of the camera in degrading images of moving objects was referred to earlier. With the freedom to implement a variety of presampling filters that is provided by a high-rate camera, the trade-off between blurring and temporal aliasing can be made on a rational basis, rather than as a happenstance of integration time. For example, in some cases it may be preferable to maintain the sharpness of moving objects even at the cost of introducing stroboscopic effects, whereas in other cases the reverse may be true. In general, temporal performance should be improved when the output of a high-rate camera is properly prefiltered before subsampling to obtain the signal to be transmitted.

Interlace

A high-rate display, for example 525 or 1050 lines progressively scanned, can positively eliminate interline flicker, even without a high-rate camera. However, if simple-minded linear interpolation is used, some loss of sharpness results. In addition, some artifacts at moving sharp edges may develop. Rather good systems have been developed in which motion-adaptive interpolation is used, temporal in the stationary portions of the image, and spatial (vertical) in the moving areas.[24] Ideally, the interpolation should be motion-compensated, i.e., along the optical flow lines, since in this way, moving objects could be kept from blurring.[25] It is also possible that fully optimized fixed interpolation, as has been developed for graphic arts applications, would serve well here.[26]

With respect to vertical motion artifacts, the disappearance of half the scan lines at certain speeds is also eliminated simply by display up-conversion, as the line interpolation process is often called. Vertical-temporal aliasing, as discussed above, is primarily a function of the camera, and can be eliminated by approprate prefiltering. Therefore, its amelioration probably requires a high-rate camera.

It should be noted that interlace

4.5.5

causes few problems in the presence of low vertical camera resolution.[g] In that case, the use of up-conversion to eliminate a nonexistent problem may, in fact, reduce rather than improve the image quality, especially if the up-conversion itself is less than perfect. The villain, in this case, is not the up-conversion, but the low resolution, which wastes channel capacity.

New Systems

The advantages and disadvantages of a revolutionary approach mirror those of the approach based on improvements in NTSC. From the long-term point of view, there is a significant economic advantage in going to an entirely new system, in addition to the fact that it would be possible to have arbitrarily high quality. Surely, if it became necessary to replace every piece of TV equipment in the world, both consumer and professional, we would all end up richer, not poorer. That route, after all, has been the history in many other fields. The creation of new industries, such as the automobile industry, is the path to wealth, not poverty. The problem is avoiding the dislocations that inevitably accompany the phasing out of products, companies, and industries.

From the technical standpoint, almost any new system can be made backward compatible by using a two-channel scheme, in which one channel is NTSC.[h] This, however, limits the efficiency of new systems. In particular, it prevents the development of any system that achieves theater quality in a 6-MHz analog channel, a possibility that we believe to be very real.

If the political obstacles can be overcome, the main advantage of a noncompatible system, in our opinion, is that its long-term economic future is much sounder than that of any compatible solution. There is no evidence that a large number of viewers will buy expensive receivers when they can see the same programs, albeit at lower quality, on their existing equipment.[i] Therefore, any compatible system runs the real risk of never developing the audience required to provide the economic justification for the production of high-definition programs on a large scale. A noncompatible system, on the other hand, requires that new receivers be used to view the new and presumably more desirable programs, especially if NTSC broadcasting is definitely to be stopped on some particular date, an-

nounced several years in advance. Since the new receivers could be made multistandard at moderate additional cost, there would also be an incentive for their purchase during the period when most of the programming were still NTSC.

From the standpoint of economic justice, this approach does not put an unreasonable burden on anyone. After all, no guarantee of permanent life is ever provided with any product, no matter how expensive or how durable. A TV receiver that cost a few hundred dollars and has lasted five years has been a much better investment than many other household appliances. We regularly trade in automobiles with a lot of life left, at a large loss, so that we can have new ones that provide only marginally better transportation. Perhaps we can learn to do the same with TV, especially as this change would probably be once in a lifetime, and the new sets could also receive the old signals.

How Good Should a New System Be?

At the very least, a new system should have a technical quality sufficiently superior to that of the current systems, as perceived by the mass of viewers in their homes, so that it would make sense to change. Some hold the view that this means at least twice the vertical and horizontal luminance resolution[j] and a higher aspect ratio. Much higher chrominance resolution is mandatory, especially in the face of higher luminance resolution. Greatly reduced spatial aliasing without attendant blurring would be highly desirable. It goes without saying that digital stereo sound is an absolute requirement of any new TV system. Although there is little discussion of temporal issues, this author would like to see smoother motion, higher spatial resolution of moving objects, and substantial absence of interline flicker. Of course, since any new system should produce the highest possible quality for the given channel capacity, it would naturally use many of the techniques discussed above for making NTSC more efficient.

It should be borne in mind that the system designer has little control over the viewing conditions. The impact of a display is greatly affected by the angle subtended at the viewer. The angle may well be different when the audience settles down to watch a spectacle on a large screen as compared to

casually observing the news on a small set over breakfast. The large-angle display requires much more spatial resolution; for pictures viewed at ten times image height, we already have enough. Thus there is some uncertainty in the following discussion.

Aspect Ratio

The motion-picture industry's great weapon of the 1950s against the inroads of television was the wide-screen film.[27] Most of the most successful films of those years were made in CinemaScope (*The Robe*, *Bridge on the River Kwai*) at 2.35:1, a format never intended to be jammed into the 4:3 "keyhole" of the small screen. Other wide-screen formats such as Cinerama (*This is Cinerama*), VistaVision (*The Ten Commandments*), and Todd-AO (*Around the World in Eighty Days*), had more fleeting popularity, but were of undoubted visual impact, as is IMAX today.[28] Two popular current formats are Panavision's anamorphic 2.35:1 (2:1 squeeze in the camera) and "flat" 1.85:1. In the face of this history, no observer tests are needed to determine the preferred aspect ratio — for entertainment productions such as drama or sports, the wider the better. For productions such as news and interview programs, as well as game and educational shows that are specifically designed for the small screen, 4:3 is just about right.

One thing that should be perfectly obvious is that thoughtful directors will turn out programs that "work" best in the intended medium.[k] If these productions are then tested in a variety of presentations, they will be preferred in the format for which they were made.[29] This presents a serious problem when material is prepared to be used alternatively in the theater or for television, as is often the case under present-day economic conditions. If a production is to look right on TV, it cannot possibly take full advantage of the wide screen of the theater unless the director takes very special pains, probably shooting certain scenes two ways. It is therefore quixotic to advocate, say, a 2.35:1 aspect ratio, which would permit much greater visual impact. However, if we are ever to break away from the 4:3 stranglehold, we should pick as wide a screen as possible, while still getting some kind of consensus to use it. In view of its popularity in the film industry, 1.85:1 is certainly a good can-

didate, and the conversion to 4:3, while difficult, is not impossible.

Transcodability

Since NTSC, PAL, SECAM, and 24 frame/sec film will be with us for a very long time, one of the most important issues in the design of any new system is the ability to be interchanged with all of these with high quality and acceptable cost. Historically, the film transcoder has proved very difficult, at least for 60-Hz systems. Since film provides much of the programming material for TV broadcasting, and the making of films by means of HDTV has been stated to be one of its important applications, this may be the most important problem.

If a TV system were to be designed only for making films, it would be perverse to use anything but 24 frames/sec, progressively scanned. This would produce films much like those shot in a film camera, but with the greater speed and convenience of electronic production. Up-conversion to 48 or 72 frames/sec for viewing would be simple.[1] It is also clear that if the audience were to be satisfied with today's motion rendition when film is shown on television, these same 24 frame/sec progressive productions could easily be converted to 60-Hz interlace (or 50-Hz interlace by 4% speedup) for broadcast in NTSC or PAL. Line-rate conversion is rather straightforward for any ratio.

A more difficult problem is presented if one wants to transcode with better motion rendition than normally seen from 24 frame/sec originals, either in the motion-picture theater or on television. Here the problem is not only transcoding but also capturing the motion adequately in the first place, a job not done very well at present by any motion-picture or TV system.

Motion Rendition[m]

Although they do not use precise words to describe what they see, nonprofessional viewers of images do notice defects in signal-to-noise ratio (SNR), sharpness, tone reproduction, and color rendition.[30] However, they are remarkably oblivious to motion defects, even of the most egregious sort. There has been very little serious study of motion perception; perhaps when there is, this situation may be clarified. It is also true that observers can be trained to notice motion problems, and when they are sufficiently

sensitized to the phenomenon (as is this author), they may regard it as an important part of overall image quality. Just as interline flicker is not apparent to lay observers until they see an interlaced display alongside one with progressive scan, it may well be that, when good motion rendition is directly compared with poor, the difference will be obvious.

In the author's opinion, good motion portrayal is a vital part of making the TV viewing experience realistic. Unfortunately, experimental work so far indicates that there is no practical frame rate that gives flawless results.[31] However, we can do much better than we do now, especially when 24 frame/sec film is used for 60-Hz TV.

Motion in Television

The main problem here involves eye tracking. If camera and eye are still, stationary objects are correctly rendered and integration time blurs moving objects for both. If the camera is tracking a moving object at which the eye is staring, again the results are fairly good, since similar blurring due to integration occurs in the eye and the camera for the portion of the scene not being tracked. In both cases, the portion of the image that is stationary on both retina and camera focal plane is rendered well. The portion in relative motion may show some stroboscopic effects, depending on the camera exposure time, but these often will not be prominent because that area is not the primary object of attention. Of course, high-rate cameras and displays, combined with the appropriate filters, help a good deal, but the rendition is fairly good even without them.

If eye and camera are not tracking in the same manner, then excessive blurring and/or stroboscopic motion are likely to take place. For example, when the eye is tracking an object moving within the TV image, it sees the blurring resulting from camera integration time. If the latter is shortened by means of a mechanical shutter or the equivalent, this blurring is reduced but the stroboscopic effect in the background is increased. Likewise, when the eye is fixed and some important object starts to move, during the period when the eye is beginning to track, very strong stroboscopic effects may occur, in the form of multiple images. These effects are not completely removed even at 120 frames/sec.[32] There appears to be no

simple way to reduce them in new systems, except to operate at as high a frame rate as possible, and to perform appropriate pre- and postfiltering in connection with high-rate cameras and displays.

Motion in Film

Typically, 24 frame/sec film is exposed with a 50% duty cycle,[n] about half the duration used in a standard TV camera. Therefore, images of moving objects will be less blurred, which is good if the object in question is being tracked within the scene by the observer, but bad if the object is also moving relative to the retina. The problems are exacerbated by the display, in which, to raise the large-area flicker rate, each frame is projected twice, or, in some projectors, three times. Now, in the tracking case, since the eye moves at the average speed of the image, double (or triple) images are clearly seen. This is more apparent with sharp-edged objects such as moving titles, but it is present nearly all the time in ordinary scenes and very obvious once pointed out. When the movie camera is tracking objects, a similar effect takes place in the background, and, depending on the subject matter, gives rise to spectacular stroboscopic effects often completely ignored by viewers.

When 24 frame/sec film is used on 60-Hz TV,[o] the difference in frame rates is smoothed over either by camera integration, which produces multiple images if these would be seen in the movie theater, or by the 3/2 pulldown method. In the latter method, successive film frames are used for two and three fields, respectively, producing a 12-Hz temporal component, and extremely jerky motion rendition.[p] As we shall see below, it is possible to do much better.

Spatiotemporal Resolution

If the conventional wisdom is followed in the choice of resolution and aspect ratio, we need about 1000 lines/frame, 1700 samples/line, and 30 frames/sec, for a total luminance data rate somewhat more than 50 Msamples/sec. This leads to bandwidth consumption on the order of that required by the NHK HDTV system. On the other hand, it is now recognized that spatial and temporal bandwidth may be traded off, reducing response for components simultaneously high frequency in all directions. In MUSE, for example, only
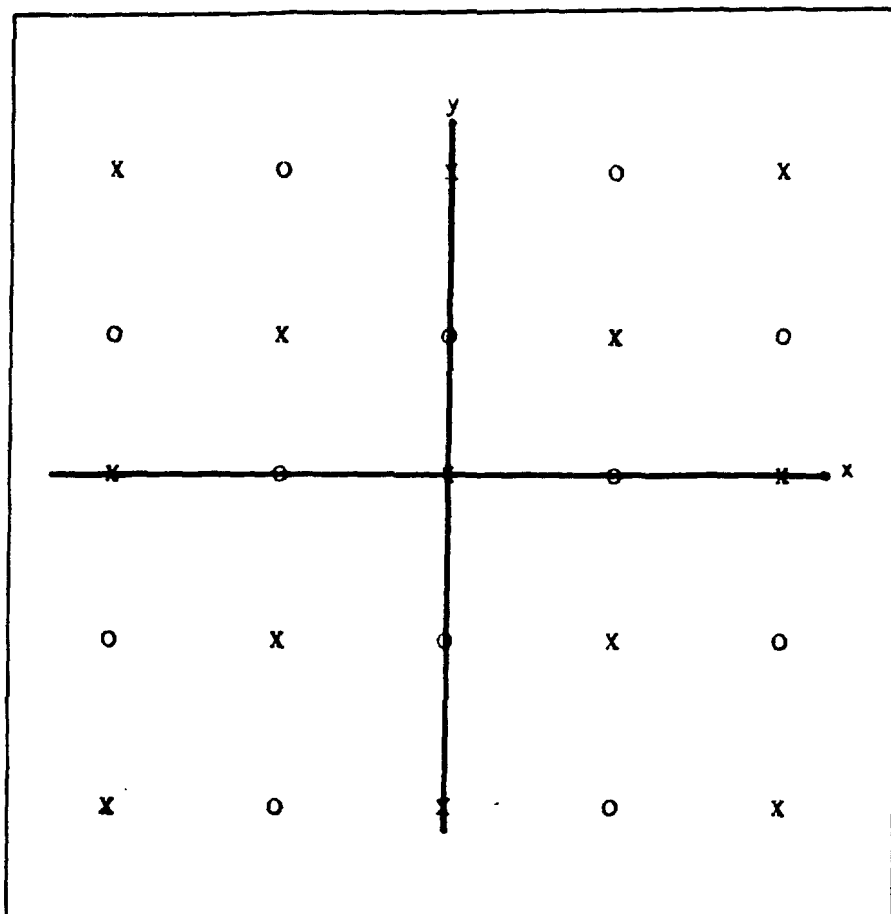
4.5.7

Figure 6. Spatiotemporally offset sampling. The Xs show the sampling grid in the even fields while the Os show the odd-field samples. With this pattern and the appropriate 3-D filters, the highest possible resolution is achieved along each axis in 3-D frequency space.

stationary objects have full spatial resolution; moving objects have only one-fourth the area resolution.[33] In Glenn's system, fine spatial detail has only one-fourth the temporal resolution.[34] While both of these methods can be thought of as compression systems, it is more meaningful to describe them as systems that attempt to use the available channel capacity in the psychophysically most advantageous manner.

One method of trading off resolution along the various axes is to use an appropriate spatiotemporal sampling pattern.[35] Each such distinct pattern leads to a distinct shape for the alias-free Nyquist baseband, which defines the resolution.[9] The offset pattern, shown in Fig. 6, makes what appears to be an appropriate compromise. Offset patterns have at least some response at frequencies 41.4% higher than the Cartesian pattern. For example, 750 lines/frame would give a peak vertical response of 1061 lines, which, with proper filtering and up-conversion, would extend the vertical MTF to more than 500 cph, or 25% higher than normally observed on the

NHK 1125-line system. Likewise, the temporal response would extend to 15 Hz with only 21 frames/sec.

Composite or Component?

As discussed earlier, we now know how to operate a composite system with essentially no cross effects. It is therefore not necessary to go to a component system, such as one of the various MAC schemes, to solve this problem.[36] In fact, the composite system has an unintended advantage, in that the portion of the 2 or 3-D spectrum that must be sacrificed to make it work does not contribute very much to luminance quality. Undeniably, however, the component schemes are simpler, which is counterbalanced, in most of those so far proposed, by the fact that they are less efficient, i.e., they take more channel capacity for the same resolution.

Since the desired shape of the Nyquist baseband can be obtained by using the appropriate offset sampling pattern together with the corresponding filters, it would appear that the better choice for new systems would be of the component type,[r] with filter-

ing and sampling to obtain a baseband in which the higher frequencies along any one axis have a lower response along the other axes. This implies that high-rate cameras and displays are used, with 3-D filtering. This, in turn, implies the presence of frame and line stores at both transmitter and receiver, so that the channel frame rate need not be high enough to avoid either large-area or interline flicker.

Sample Systems

We will now briefly discuss a few sample systems. None of these is seriously proposed as a standard of any kind; each is mentioned only for the purpose of illustrating specific points. All new systems are to use frame stores and high-rate cameras and displays. The question of aspect ratio is finessed by speaking only of the total number of samples per frame; these may be rearranged as desired.

A System for the 6-MHz Analog Channel.

Assuming that it would be harder to change channel allocations than TV system standards, it is of interest to see just how good an image can be transmitted in the existing terrestrial channels. We note that only a 4.2-MHz signal can at present be used in a 6-MHz channel, ~~a system like that used in AM stereo~~, and that the horizontal definition is reduced by more than 15% by retrace time. The first step, therefore, is to eliminate vestigial-sideband transmission and the separate sound carrier, and either to eliminate the retrace time or use it for something important. A single carrier, in the middle of the band, could carry two independent 3-MHz channels by double-sideband quadrature modulation, a technique like that used in AM stereo. To minimize adjacent-channel interference, the outer extremities of the band should carry only detail information (rather than any low-frequency data), perhaps by having each of the two signals carry information for half the lines.

Assuming that some advantage in motion portrayal will ensue from use of proper filtering, we should be able to use no more than 18 frames/sec.[5] If the channel, now a full 6 MHz, is good for 12 Msamples/sec, we can have 667,000 samples/frame, or at least three times that used at present. We still must provide for color and audio. Sixteen percent of the capacity should

be sufficient for the former and 8% for the latter, still leaving a picture of about 500,000 samples/frame — very much better than the 100,000 or so typically seen on home receivers at present.

A System for 216 Mbits/sec

CCIR 601, the standard for digital transmission of current TV signals, provides an image substantially superior to that of the best analog transmission, even without using any special advanced techniques. Since transmission facilities and, perhaps more importantly, digital VTRs are being provided for this standard, it is interesting to see what can be done to raise image quality using some of the methods mentioned here.

Digital transmission permits the use of two additional techniques: spatially offset sampling as advocated by Wendland,[37] and relatively coarse quantization of samples representing the spatial high frequencies.[38] In the latter scheme, each frame is divided by 2-D spatial filters into highs and lows.[1] The three-color lows are coarsely sampled and finely quantized, and the monochrome highs are finely sampled and adaptively coarsely quantized. Using 24 frames/sec and allowing a full 2.5 Mbits/sec for digital audio gives nearly 9 Mbits/frame. If the lows have one-third resolution in each direction, each block of 9 samples requires 27 bits for the color lows using 9 bits/sample, giving a SNR in the blank areas of 59 dB. The highs take 4 bits/sample plus 4 bits/block of adaptation information for a total of 7.44 bits/sample, allowing nearly 1.2 Msamples/frame. When these are spatially offset in successive fields as shown in Fig. 6, and then interpolated up to double resolution for display at 72 frames/sec, progressively scanned, we would expect to have images superior to those of the NHK HDTV system with respect to flicker, motion smoothness, sharpness of moving objects, spatial resolution, and SNR.

A Very High Quality TV System (VHQTV)

Hitachi has made an experimental 500 Mbit/sec DVTR and Sony is said to have made a similar machine of about 1 Gbit/sec capacity. If such were readily available, with the cameras and displays to match, what kind of a system would be most appropriate? It is hard to say. Surely, the 216

Mbit/sec system described in the preceding section would satisfy the most demanding consumer application. However, there may be a place for even higher performance for certain kinds of public displays. Two such systems have been shown in the theater. One is IMAX, 24 frames/sec with a very large film area (48×69 mm), and the other is Showscan, 60 frames/sec with a more modest film area, but still much larger than standard 35mm. Without extensive study, it would not be possible to determine the best way to use the additional capacity. To hedge the bet, one could increase the frame rate modestly, to 36 frames/sec. What to do about the spatial rendition depends largely on whether the additional 50% of spatial data is to be used to increase the sharpness or the size of the final image. If the sharpness is to be increased, then adding a third channel, at perhaps 2 bits/sample, would double the area resolution. If the image size is to be increased, then it would go up only in proportion to the data rate — 50% on an area basis.

Motion Compensation

It has been known for a long time that successive frames must be very much alike to achieve an illusion of continuous motion. This has been the basis of a number of coding and noise-reduction proposals.[39] A more advanced concept is to take advantage of knowledge of the actual motion, or displacement, of corresponding points in successive frames. Such knowledge can be used for noise reduction without blurring, predictive coding, and adaptive interpolation. Like other adaptive techniques, motion-compensated processing is always better than nonadaptive processing, which is, after all, just a special case, *except when the motion estimation is wrong.*[u] Present-day estimation methods are good enough for important practical applications. The major obstacle is not the accuracy, but the speed and cost of the associated calculations.

Evidence that motion compensation is useful is given by three recent demonstrations. NHK uses motion-compensated interpolation for MUSE[40] and for the HDTV-to-PAL transcoder.[41] The BBC Research Laboratories has shown a 4:1 compression system, called digitally assisted TV (DATV), using subsampling and motion-compensated interpolation.[42] The first two systems,

which are quite successful, use a rather crude form of this processing. DATV works well where the motion estimates are accurate.

Motion Estimation at Receiver or Transmitter?

An important issue is whether the motion information is to be calculated at the transmitter and sent in a side channel or whether it is to be calculated at the receiver. Using a side channel for auxiliary data effectively broadens the signal description to include velocity information. There is no principle of signal processing that shows that the optimum signal description consists only of a sequence of sample values. In describing the focal plane light intensity function (which we prefer to call the *video function*, to distinguish it from the *video signal*), it is quite possible that a more efficient (compact) description would include some motion information, which corresponds, more or less, to derivative information. In addition, if a high-rate camera is used, there is more data available at the transmitter to calculate the motion accurately. Finally, since there is one transmitter and many receivers, good system design would put the major burden of this calculation at the sending end. It is possible that a simple calculation could be done at the receiver, with the side channel being used to transmit supplementary information.

Motion-Compensated Temporal Interpolation

An important application of this technique is to get a more advantageous trade-off between spatial and temporal resolution. For example, in the 24 frame/sec systems described previously, comparable performance could very likely be obtained at 12 frames/sec. For the same data rate, this would permit doubling the area spatial resolution, for a notable improvement in overall image quality.

Another application is the conversion of 24 frame/sec film to 60-Hz TV. We have done some experiments on this subject in our laboratory with promising results. Especially with computer-generated images having the full spatial resolution permitted by the standards, the elimination of jerky motion is quite spectacular. With natural (camera) images, the improvement, although noticeable, is much less, a fact we attribute mainly to typically poor camera performance.

4.5.9

The performance at 24 frames/sec was so good, we have tried one experiment at 12 frame/sec, using only motion information derived from the final 12 frame/sec signal. (In a practical case, we would have the original camera sequence available and could do a better job of motion estimation.) The results were, at the very least, promising. There is a strong possibility, in our opinion, that 12 frames/sec will be shown to be a completely adequate frame rate when motion compensation is used for final viewing.

Motion-Compensated Noise Reduction

Noise reduction is of commercial importance. Although it is true that modern TV cameras, with adequate illumination, produce signals that need no noise reduction, multigeneration analog recording, ENG equipment, low-light-level picture taking, and satellite transmission all produce signals that have greater-than-optimum noise levels. The present generation of frame-recursive motion-adaptive noise reducers, all descendants of ideas proposed in the early 1970s,[43] trade off noise reduction against blurring of moving objects. The effective time constant of a temporal low-pass filter, using feedback around a single-frame memory, is adjusted in accordance with the amount of frame-to-frame motion, imputed from the measured difference between the same points in successive frames. A great deal of smoothing is used in stationary areas and less and less in the presence of more and more motion. In some cases, a "slow switch" in used, in which the time constant is based on an area measurement rather than on a point-by-point measurement.

These noise reducers tend to blur out low-contrast moving detail and they fail to reduce noise in moving objects. When the latter are tracked by the viewer, the noise visibility is hardly reduced. Both of these defects can be cured by integrating in x,y,t space, not parallel to the time axis, but along the motion vector. If smooth motion persists long enough to be tracked (more than about 200 msec), it would appear that the noise could be reduced at least 6 dB, and in most cases, 12 dB or more, with negligible reduction in sharpness.[44]

Conclusion

After a discussion of the limits to

received image quality in NTSC and a review of various proposals for improvement, it is concluded that the current system is capable of significant increase in spatial and temporal resolution, and that most of these improvements can be made in a compatible manner. Newly designed systems, for the sake of maximum utilization of channel capacity, should use many of the techniques proposed for improving NTSC, such as high-rate cameras and displays, but should use the component, rather than composite, technique for color multiplexing. A preference is expressed for noncompatible new systems, both for increased design flexibility and on the basis of likely consumer behavior. Some sample systems are described that achieve very high quality in the present channels, full "HDTV" at the CCIR rate of 216 Mbits/sec, or "better-than-35mm" at about 500 Mbits/sec. Possibilities for even higher efficiency using motion compensation are described.

Acknowledgments

Appendix — The Spectrum of the Composite Signal

We call the light intensity in the focal plane of a camera the video function and note that it is a function of $x, y,$ and $z,$ with z standing for time. Within the borders of the image and within an arbitrary time interval, the function can be represented by a triple Fourier series. The latter, being periodic in the width $A,$ height $B,$ and time interval $C,$ thus replicates a triple infinity of identical functions outside the A,B,C box.

A typical term of the Fourier series is

$$\exp[2\pi j(mx/A + ny/B + pz/C)]$$

where $m, n,$ and p are the number of cycles per image width, height, and duration, respectively. We call these terms spatiotemporal harmonics. To find the signal that results from scanning, we note that repetitive scanning downward within the box is equivalent to unidirectional scanning in x,y,z space according to the expressions

$$x = Af_h t; \; y = -Bf_v t; \; z = Cf_v t$$

Thus, the video signal is represented by a temporal Fourier series with a typical term

$$\exp[2\pi j(nf_h - mf_v + pf_v)t]$$

where each spatiotemporal harmonic of the video function corresponds to one temporal harmonic of the video signal.

In the NTSC system, the color subcarrier frequency is related to the horizontal and vertical scanning frequencies by the following expressions:

$$f_{sc} = (455/2)f_h = (455/2)(525/2)f_v$$

The positive frequency color subcarrier can thus be thought of as resulting from the scanning of a spatiotemporal harmonic in which

$$n = 227, m = -131, f_v = 15 \text{ (Hz), and}$$
$$n = 228, m = 131, f_v = -15$$

while the negative subcarrier frequency results from

$$n = -227, m = 131, f_v = -15, \text{ and}$$
$$n = -228, m = -131, f_v = 15$$

Note that for a harmonic to be real, there must be a pair of exponential (complex) harmonics in diametrically opposite octants.

The subcarrier harmonics are located in four of the eight octants in 3-D frequency space, as shown in Fig. 1. Each such harmonic is surrounded by sidebands representing the chrominance signals. Since the scanning patterns for chrominance and luminance are the same, the "natural" bandwidths in all three directions are also the same. The horizontal chrominance bandwidth is restricted by the temporal filters used before and after combining the components to form the composite signal, but there are normally no provisions for limiting the vertical and temporal bandwidths. As shown in Figs. 2 and 3, if cross effects are to be controlled, it is required to shape both the luminance and chrominance spectra in all three dimensions. Most proposed techniques result in substantially less bandwidth for chrominance than for luminance, as well as reduced spatial resolution at higher temporal frequencies for all components.

References

1. W. F. Schreiber. "Psychophysics and the Improvement of Television Quality," *SMPTE J.*, 93:717-725, Aug. 1984.

4.5.10

2. D. Lemay, "Interpolation Temporale avec Compensation du Mouvement," M. Sc. Thesis, INRS-Telecommunications, May 1985.
3. A. Roberts, "The Improved Display of 625-Line Television Pictures: Adaptive Interpolation,: BBC Res. Report., May 1985.
   J. L. E. Baldwin, "Enhanced TV — A Progressive Experience," SMPTE J., 94:904-913, Sept. 1985.
   K. H. Powers, "Techniques for Increasing the Picture Quality of NTSC Transmission in Direct Satellite Broadcasting," IEEE J. Sel. Areas Comm., 3:57-64, Jan. 1985.
4. R. Turner, "Some Thoughts on Using Comb Filters," IEEE Trans. Consumer Electr., CE-23, 3:248-256, Aug. 1977. S. J. Auty et al., "PAL Color Picture Improvement of Television Quality," SMPTE J., 87:677-681, Oct. 1978.
5. T. Fujio et al., "HDTV," NHK Tech. Monograph, 32, June 1982.
6. J. L. E. Baldwin, "Analog Components, Multiplexed Components, and Digital Components — Friends or Foes?," SMPTE J., 92:1280-1286, Dec. 1983.
7. E. F. Brown, "Low-Resolution TV: Subjective Comparison of Interlaced and Non-Interlaced Pictures," Bell Sys. Tech. J., 46:199-232, Jan. 1967.
8. T. G. Schut, "Resolution Measurements in Camera Tubes," SMPTE J., 92:1270-1293, Dec. 1983.
9. I. C. Abrahams, "The Frequency-Interleaving Principle in the NTSC Standards," Proc. IRE, 42:81-83, Jan. 1954.
10. G. T. Flesher et al., "The General Theory of Comb Filters," Proc. NEC, 17:81-94, 1961.
11. F. Gray, "Electro-optical Transmission System," U.S. Patent No. 1,769,920, Apr. 30, 1929.
12. I. C. Abrahams, "Choice of Chrominance Subcarrier Frequency in NTSC," Proc. IRE, 41:79-81, Jan. 1954.
13. M. W. Baldwin, "The Subjective Sharpness of Simulated TV Pictures," Proc. IRE, 458-468, Oct. 1940.
14. J. Fourier, "Theorie Analytique de la Chaleur," Didot, Paris, 1822.
15. P. Mertz and F. Gray, "A Theory of Scanning and its Relation to the Characteristics of the Transmitted Signal in Telephotography and Television," Bell Sys. Tech. J., 13:464-515, July 1934.
16. N. W. Parker, "Color TV Signal Separation System," U.S. Patent No. 3,542,945, Sept. 11, 1967.
17. J. O. Drewery, "The Filtering of Luminance and Chrominance to Avoid Cross-Colour in a PAL Colour System," BBC Engineering, 8:8-39, Sept. 1976.
   E. Dubois et al., "The Three-Dimensional Spectrum and Processing of Digital NTSC Color Signals," SMPTE J., 91:372-378, Apr. 1982.
18. E. Dubois and P. Faubert, "Two-Dimensional Filters for NTSC Color Encoding and Decoding," IBC 86.
19. Y. Faroudja, "Adaptive Comb Filtering," U.S. Patent No. 4,179,705, Mar. 13, 1978. J. Rossi, "Comb Filter for TV Signals," U.S. Patent No. 4,050,084, July 14, 1976.
   A. Kaiser, "Comb Filter Improvement with Spurious Chroma Deletion," SMPTE J., 86:1-5, Jan. 1977.
20. T. Fukinuki and Y. Hirano, "Extended Definition TV Fully Compatible with Existing Standards," IEEE Trans. Comm., 32:948-953, Aug. 1984.
21. T. Fukinuki, Y. Hirano, and H. Yoshigi, "Experiments on Proposed Extended-Definition TV with Full NTSC Compatability, SMPTE J., 93:923-929, Oct. 84.
22. B. Wendland and H. Schroder, "Signal Processing for New HQTV Systems," SMPTE J., 94:182-189, Feb. 1985.
23. S. Hsu, "The Kell Factor: Past and Present," SMPTE J., 95:206-214, Feb. 86.
24. R. N. Jackson and M. J. J. C. Annegarn, "Compatible Systems for High-Quality Television," SMPTE J., 92:719-723, July 83.
25. D. M. Martinez, "Model-Based Motion Estimation and its Application to Restoration and Interpolation of Motion Pictures," Ph.D. Thesis, M.I.T. EECS Dept., Aug. 1986.
26. W. F. Schreiber and D. E. Troxel, "Transforma-

tion Between Continuous and Discrete Representations of Images: A Perceptual Approach," IEEE Trans. Pat. Anal. & Mach. Intel., 7, Mar. 1985.
27. "Widescreen," Special issue of The Velvet Light Trap, 21: Summer 1985, Madison, Wis.
28. W. C. Shaw et al., "IMAX and OMNIMAX Theatre Design," SMPTE J., 92:284-290, Mar. 1983.
29. H. D. MacDonald, "The Relative Quality of Television and Motion Pictures, M.S. Thesis, M.I.T. EECS Dept., Jan. 1985.
30. R. W. Neuman et al., "Report of Preliminary Focus Group Interviews," MIT/ATRP Report, 1984, unpublished.
31. G. Tonge, "Les Assises des Jeunes Chercheurs," Meeting Report, Rennes, France, Sept. 1985.
32. S. C. Hsu, "Motion-Induced Degradations of Temporally Sampled Images," M.S. Thesis, M.I.T. EECS Dept., June 1985.
33. Y. Ninomiya, Y. Ohtsuka, and Y. Izumi, "A Single Channel HDTV Broadcast System — The Muse," NHK Lab. Note 304, pp. 1-12, Sept. 1984.
34. W. E. Glenn, "Reduced Bandwidth Requirements for Compatible HDTV Transmission," Presented at IGC HDTV Conf., New Orleans, Apr. 1984.
35. Op cit Ref 1.
36. D. Lowry, "B-MAC, An Optimum Format for Satellite TV Transmissions, "SMPTE J., 93: 1034-1043, Nov. 1984.
   H. Mertens and D. Wood, "The C-MAC Packet System for Direct Satellite TV," EBU Tech. Rev., 200: Aug. 1983; also IEEE J. Sel. Areas in Comm., SAC-3, 1:65-75, Jan. 1985.
37. Op cit Ref. 22.
38. W. F. Schreiber and R. R. Buckley, "A Two-Channel Picture Coding System: II-Adaptive Companding and Color Coding," IEEE Trans. on Comm., Com-29, 12:1849-1858, Dec. 1981.
39. J. E. Cunningham, "Temporal Filtering of Motion Pictures," Ph.D. Thesis, M.I.T. EE Dept., June 1964.
   F. W. Mounts, "A Video Coding System with Conditional Replenishment," Bell Sys. Tech. J., 48, Sept. 1969.
40. Op cit Ref. 33.
41. M. Sugimoto et al., "Second Generation HDTV Standards Convertor," 14th Int'l. TV Symposium, Montreux.
42. R. Storey, "HDTV Motion Adaptive Bandwidth Reduction Using DATV," BBC Research Dept. Report BBC RD 1986/5.
43. D. Connor, Apparatus for Improving Video SNR," U.S. Patent No. 4,040,221, Mar. 13, 1972.
   J. Drewery and M. Weston, "Noise Reduction in TV Signals,: U.S. Patent No. 4,058,836, Apr. 26, 1976.
   A. Kaiser, "Noise Reduction System for Color TV," U.S. Patent No. 4,064,530, Oct. 10, 1976.
44. E. Dubois and S. Sabri, "Noise Reduction in Image Sequences Using Motion-Compensated Temporal Filtering," IEEE Trans. Comm., 73:502-522, July 1984.

Bibliography

D. G. Fink, "The Future of High-Definition Television: First Portion of a Report of the SMPTE Study Group on High-Definition Television," SMPTE J., 89:89-94, Feb. 1980.
D. G. Fink, "The Future of High-Definition Television: Conclusion of a Report of the SMPTE Study Group on High-Definition Television," SMPTE J., 89:153-161, Mar. 1980.
G. J. Tonge, "The Sampling of Television Images," IBA Experimental and Development Report 112/81.
G. J. Tonge, "Signal Processing for HDTV," IBA Report E8D, July 1983, U.S.A.
D. E. Pearson, Transmission and Display of Visual Information, Pentech Press, 1975.

Notes

a. Most of what is said here about NTSC applies to PAL and SECAM as well.

b. Another way of putting this is that if interline flicker is not observed, the camera has insufficient vertical resolution and is therefore wasteful of channel capacity.

c. It also limits the alias-free bandwidth to ±7.5 Hz, rather than ±15 Hz, as in the monochrome system.

d. This material is taken from a forthcoming tutorial paper by Eric Dubois and the author.

e. Note that all transversal filters are "comb" filters, i.e., they have a periodic temporal frequency response. The usefulness of comb filters in separating luminance and chrominance, which have interleaved periodic spectra, is obvious. However, efficient separation requires using a rather large number of line delays. The usual one or two-line comb filter does not separate the signals very well, without at the same time substantially reducing vertical resolution, at least at the higher horizontal frequencies.

f. The use of this term is associated with the zone plate test pattern. Such components also result from two-dimensional texture.

g. If the screen is viewed closely enough so that the scan lines are clearly resolved, interline flicker will be seen even with a blank field, a problem that requires display up-conversion for correction.

h. A somewhat tongue-in-cheek proposal is to have one NTSC channel and one PAL channel, the combination becoming HDTV! With advanced signal processing and some tolerance of reduced performance on existing receivers, this is not entirely out of the question.

i. There is, on the other hand, copious evidence that viewers are willing to pay substantial amounts of money for programs that they want to see, even when many other programs are available free.

j. This specification is indefinite, since it is not apparent whether the comparison is to the vertical resolution achieved with NTSC on standard receivers or on properly up-converted displays, with signals originating in a high-line-rate camera. See the Fink articles mentioned in the bibliography.

k. In our Audience Research Facility, we carried out one test of preferences for aspect ratio in monochrome photographs. Not surprisingly, portraits were preferred in "portrait" mode and landscape in "landscape" mode. I have no doubt that this is true for TV as well.

l. Editing would probably be even easier with a frame synchronizer, fed directly by the 24 frame/sec signal, but supplying 525-line, 60-Hz interlaced signals to the monitor.

m. This section is based on discussions with Andrew Lippman and Stephen Hsu.

n. This is also referred to as a 180° shutter.

o. In 50-Hz countries, the film is either shot at 25 frames/sec rather than 24 frames/sec, or is sped up by 4%. Motion rendition is similar to that seen in the theater.

p. This author finds it striking that this is tolerated by American TV professionals. In contrast, the insistence of European TV interests on near-perfect rendition in the NHK HDTV-to-PAL transcoder seems to imply a quite different view of TV quality.

q. These basebands are not unique. However, there is a "natural" baseband, of maximum symmetry, for each sampling pattern. Except for the Cartesian pattern, they all have a temporal frequency response that varies continuously with spatial frequency, and vice versa.

r. It follows that the composite technique should find its principal application in improving existing systems, and not in the design of new ones.

s. Eighteen frames and 36 fields/sec, even with simple-minded temporal interpolation up to the display rate, should give better motion smoothness than in motion pictures. The latter is acceptable to audiences around the world, even though many defects can be seen by experts.

t. There is a 3-D extension of this system, which is somewhat more efficient, but its quality has not yet been completely proved out.

u. Noise is the enemy of accurate motion estimation. Recursive estimation, in which the noise is reduced, permitting more accurate estimation as the estimate converges, shows some promise.

TAB

Television's Search for Tomorrow

William F. Schreiber

This article, which was intended for publication in MIT's *Technology Review,* was never published because of an editing dispute. It was widely distributed in October 1985 and has been read by many people, here and abroad, engaged in television research.

In this paper, I have attempted to make the technological issues pertaining to advanced television systems understandable to intelligent nonscientists. I have also dealt with the international regulatory controversy that has developed as a result of the attempt by the US and Japan to have the Japanese (NHK) HDTV system adopted as an international legal standard for television production and program interchange. I pointed out the extraordinary progress being made in the development of new and more efficient TV systems, and ended with a plea against premature adoption of what I believed (and still believe) to be a poorly conceived system.

Since writing this paper, many more alternative systems have been proposed, and some have been demonstrated. I believe the adoption of the NHK system as a production and exchange standard is against the best interests of the United States. There is no pressing need to adopt anything at the present time. We can afford to wait for something more in accord with American interests, both consumers and industry.

This is a draft of a paper scheduled to be published in *Technology Review* for January 1986. Some minor changes will be made in the final version. Unfortunately, the computer-generated figures that illustrate many of the points made in the paper have not yet been completed.

Television's Search for Tomorrow
by
William F. Schreiber

William F. Schreiber is Professor of Electrical Engineering at MIT and Director of the Advanced Television Research Program. His major professional interest always has been image processing, both theoretical and applied. The opinions expressed herein are those of the author and not of MIT or the sponsors of the Advanced Television Research Program. In a number of cases, they are at variance with sponsors' stated positions.

Premature adoption of an international technical standard for television production could prejudice attempts to develop greatly improved television systems for the future.

There is a strong possibility that the television system that has had such a profound effect on our culture will, within the next five or ten years, change in a number of significant ways. If we so desire, we can have a system with larger, sharper pictures, equal to or better than 35mm theater film. We can have better rendition of motion, improved sound, and worldwide compatibility. If the quality were high enough, an entire new industry -- electronic still photography -- might be created.

Building and selling all the equipment needed for such new systems would give an important stimulus to the world economy. US industry, if it moved fast enough and proved smart enough, might get a good part of the market. However, this alluring future may be derailed by what some, including myself, believe to be the premature adoption of a proposal for a production standard for high-definition TV.

The proposal, which originated in Japan and is supported by some American organizations as well, calls for programs intended for international exchange to be produced, although not transmitted, using the standards for a high-definition TV system developed under the leadership of Japan Broadcasting Corporation (NHK). The State Department has adopted the proposal as US policy, and supported it in October at the quadrennial standards-setting meeting of the International Radio Consultative Commission (CCIR). Opponents of the standard, particularly among European broadcasters, argue that it would cost a great deal to adopt as long as present broadcasting standards are in use, and that it fails to recognize current rapid progress in TV technology. By accepting the standards now, critics argue, the chances for future innovation are likely to be stifled.

It should be understood that I am talking about the "medium" and not the "message." Of course a TV system that simulated looking at the real world through a wide clear window, rather than through a gauze-covered keyhole, would permit, perhaps even stimulate, a different style of programming. However, it is clear that technology alone cannot correct any of the perceived shortcomings of today's TV programs. Thus to some, efforts to improve the technical quality of the image might seem beside the point. However, since TV is an economically important industry, and one in which the American share has fallen in recent years, the potential for a renaissance through better technology is quite appealing. In addition, working with images is aesthetically satisfying and intellectually challenging; as in many similar cases, developments in the field are likely to be driven to some extent by the engineers and scientists who are attracted to the field.

Why Now?

Our present television system has proven remarkably durable, especially considering the radical changes in technology that have taken place since its adoption more than thirty years ago. However, a number of contemporary events and developments have combined to provide a setting in which change has become more likely.

Alternative transmission channels. Television is a voracious consumer of bandwidth, or channel capacity, one present-day TV channel using four times as much of this limited resource as the entire AM radio band. Many of the proposed improved systems require even more bandwidth, and so cannot be accommodated, with today's technology, in the existing 6 MHz channels. This additional capacity could be provided by cable, which has more than it needs, by direct broadcasting to the home from satellites (DBS), or by improved videocassettes and videodiscs. Cable is already here, DBS can be provided as soon as economically feasible, and magnetic recording is rapidly improving. Many fiberoptics cables have

already been laid for common carrier communication systems, and could be run into every household i the "wired city" ever became a reality.

Semiconductor development. The sophisticated TV systems of the future will require complex signal processing and a substantial amount of computer-like memory in the receiver. This will be practical only because of the revolution that has been sweeping the semiconductor industry in recent years. The cheaper and more powerful chips that set the stage for microcomputer development also provide the possibility of vastly increasing the "intelligence" of receivers, a necessity if improved image quality is not to require greatly increased bandwidth. For example, if we wanted to improve motion portrayal by doubling the number of separate pictures transmitted per second, that would double the bandwidth. On the other hand, a receiver with enough computer power might be able to generate these extra frames from those already transmitted, without raising the bandwidth consumption at all.

Signal processing. After forty years' effort, we have learned a great deal about processing the signals involved in imaging systems. For example, the ability of Caltech's Jet Propulsion Laboratory to produce spectacular improvement in the quality of photos from space is well known. Virtually every excellent color image we see in a magazine has been digitized and computer-processed for enhancement of quality. Comparable processing of the moving images of television and motion pictures in real time is more challenging, but research workers are making notable progress. The Bell Northern Research Laboratory in Montreal, for example, has demonstrated, by computer simulation, very high quality color images for video conferencing using only *one fiftieth* the normal channel capacity. Image sequences with only slightly restricted motion capability have been produced using only every eighth original frame. A number of laboratories around the world have produced significantly improved image quality from standard broadcast signals by means of signal processing in conjunction with special receivers.

A significant point is that all of this processing is digital. While the existing TV industry uses mostly analog techniques, a certain amount of digital processing has already found its way into TV studios. Many of the elaborate special effects, including all that require continuously variable rotation or change of size, are done digitally. Network coverage of the most recent Olympics and of the 1984 national elections made extensive use of so-called digital video effects (DVE) systems for this purpose. A number o companies have demonstrated digital video tape recorders, although they are not yet in commercial use. Even now, however, virtually every professional video tape recorder (VTR) includes a digital time base corrector that cancels out the effect of nonconstant head and tape speeds.

Le defi Japonais. No doubt, the most compelling impetus toward change was the demonstration, in

1981, of the Japanese high-definition television system (HDTV), after a development that took more than ten years. It was orchestrated by NHK, which coordinated the work of many other laboratories, including Hitachi, Matsushita, and Sony. The results are roughly the equivalent of 35mm motion pictures and are quite striking to anyone accustomed to normal home TV. This significant development, which could not have happened in the US, (for one thing, its organization would probably have been illegal under our antitrust laws) includes high-definition cameras and picture tubes as well as an analog VTR, film scanner, film recorder, and associated circuitry. The cameras, picture tubes, and VTR represent a substantial advancement in the state of the art, and are therefore the subject of universal admiration.

The system tying all these new components together, however, is quite conventional; thus it requires four to five times the channel capacity of existing systems. This requirement, coupled with the noncompatibility of the system with existing receivers, is at the root of much of the controversy. A subsampled version, the ingenious MUSE system, requires a more sophisticated receiver but only about twice the present bandwidth, and it is also noncompatible. The image quality of MUSE has not been fully proved. While it is much higher than NTSC, it almost surely is lower than that of the original HDTV system.

How TV Works, and How it Sometimes Fails

Scanning. Production of a television signal involves a series of abstractions of the real, three-dimensional world in front of the camera. Light reflected from the scene forms an image in the focal plane of the camera. [SEE BOX] The point-by-point image intensity is sampled by an aperture (physically, an electron beam) that repeatedly scans the focal plane in a regular pattern of closely spaced parallel lines, called a raster. The video signal transmitted to the receiver is proportional to the image intensity sensed by the aperture. At the receiver, a scanning electron beam in the picture tube, synchronized with that at the camera (but delayed, if need be, by the transmission time) traverses the phosphor-coated receiving screen in the same raster pattern. The instantaneous beam intensity is made proportional to the light intensity on the focal plane of the camera, producing the output image by translating the energy of the beam into light.

The pattern of light emitted from the screen depends on the persistence of the phosphor. To avoid carryover of light from one frame to the next, which would blur moving objects, the persistence is always less than a frame time, and usually very much less — typically less than one half. As a result, the instantaneous screen image actually comprises a small number of very intense scanning lines, moving down the screen. The perception of a continuously illuminated image depends entirely on integration in the eye, in

all present systems.

The image quality obviously depends on the number of complete pictures per second, the number o lines per picture, and the number of resolvable picture elements (pels or pixels) per line. Higher numbers give higher quality but naturally demand higher channel capacity. The required bandwidth of the transmission channel, in Hertz (cycles per second), is about one half the number of pels per second. Thus a 30x30 image, transmitted at ten frames per second, requires 9000 samples per second or 4500 Hz, which is roughly the bandwidth of a radio voice channel. In fact, the great Scottish TV pioneer, James Logie Baird, transmitted such pictures from ships at sea, in color and in 3-d, and at night with infrared illumination, in the 1920's. Of course, the quality of such low resolution images was rather poor.

As time went on, and the relationship of the scanning parameters to the channel capacity became known (Baird was innocent of such matters), image quality steadily improved. Britain used a 405-line, 25 frame per sec (fps) system for the coronation of King George VI in 1937. The French used an 819-line system for some time after World War II. All of Europe went to 625 lines, 25 fps in 1965(?) at the time color was added. An earlier experimental 441-line, 30-fps monochrome system was supplanted by the first American commercial system in 1941, using 525 lines, 30 fps. By comparison, the NHK HDTV system uses 1125 lines, although barely 800 are actually resolved on the screen. This inefficiency of utilization of vertical resolution is common to all existing systems, but can be mostly eliminated by the signal-processing techniques previously mentioned.

The number of pels per frame is a compromise between bandwidth and resolution, or sharpness. The 525-line NTSC compromise yields images with spatial resolution far too low for the large screens we take for granted in the theater. For comparison, office facsimile equipment normally uses about 1000 lines per page with equal horizontal and vertical resolution. Xerox copies are equivalent to about 2000 to 3000 lines, as are good quality printed color images. 35mm motion picture film is equivalent to 1000 to 1500 lines, and amateur slides to 2000 to 3000 lines.

Another difficulty with upgrading TV to theater quality is the aspect ratio – the ratio of width to height of the screen. Current TV systems have a 4:3 aspect ratio because that was standard for film in 1936 when TV parameters were first chosen. Most films are now made at 1.85:1 or 2.35:1 and it is likely that any new TV system would have a wider picture, such as NHK's 5:3. There is no question but that a large screen and a very wide field of view add greatly to the sense of realism. In fact, systems such as Cinerama and Todd-AO *(Around the World in 80 Days)* achieved a good sense of three-dimensionality using neither stereo nor polarizing glasses.

The number of frames per second is a compromise between bandwidth on the one hand and motion rendition and flicker on the other. The 25- and 30-fps systems use that frame rate, in the countries that have 50- and 60-Hz power systems respectively, because there once was an advantage in having it nearly identical to a (sub)multiple of the power frequency. Motion pictures use 24 fps. (In about 1928, the rate was raised from the 16 fps of silent movies for the sake of the *sound* quality!)

Since the eye's ability to integrate the wildly fluctuating light from the screen is limited, these rates, i used in TV and motion pictures, all would cause unacceptable flicker. Therefore the display rate must be raised. In film, each frame is repeated two or, more commonly, three times. Repetition is not possible in simple TV systems, so another technique — interlace — is used. In all current TV systems, alternate lines of the raster are transmitted on alternate scans of the screen. Thus, in the US system, 60 images (called fields) are displayed each second, each frame having 262 1/2 lines. This gives a large-area flicker rate of 60 per second, almost eliminating the flicker, provided the screen is neither too bright nor too wide. Viewers see more flicker in the 25-fps countries, but generally ignore it. The small-area, line-to-line flicker rate is still 30 per second, however, which is intolerable if the lines can be clearly resolved, as when viewing from close up. For this reason, interlace is usually not used in video display terminals. Flicker is more perceptible at the periphery of the visual field, so that it is often evident at the edges o wide-screen movies, even at 72 flashes per second, when we have our gaze fixed at the center. (It generally disappears if we look directly at the edge of the screen.)

Owing to certain details of the operation of interlace in TV cameras, the picture rate that relates to the rendition of motion is actually the field rate and not the frame rate. Thus, existing TV systems have adequate, although not perfect, motion rendition, rapidly moving objects tending to become blurred, the effective exposure time being 1/60 sec. ABC's "Super Motion" system, developed by Sony, effectively introduced a 1/180 sec shutter into the camera in a partially successful attempt to make the temporal resolution even higher. Flicker at the periphery, which we noted is present in wide-screen film even at 72 flashes per sec., is by itself a barrier to really large TV pictures at 60 fields per sec. In fact, movies portray motion far less smoothly and accurately than TV. Surprisingly, most viewers seem to ignore this problem. Possibly they unconsciously accept it as part of the "film look."

Interlace has problems as well as benefits. The effective vertical resolution is reduced from 525 lines to something less — no more than 75% and under some conditions only 50%. This is partly because o the physics of camera operation and partly for psychophysical reasons. In any event, if the camera and display had the full 525 lines of resolution called for by the standards, the interline flicker would be much worse. In addition, vertical motion of the camera produces certain artifacts, including the disappearance

of half the lines, a phenomenon that is easy to observe by moving one's finger slowly down the face o the tube. Recently developed noninterlaced systems (called progressive scanning and requiring twice the bandwidth) show a remarkable improvement in "quietness" or "stability" when displayed side-by-side with standard systems. Although we have become accustomed to the shimmery nature of interlaced TV, we can easily see and appreciate the difference.

Color. [SEE BOX] Color television requires the transmission of three separate images, made through three different color filters, thus representing three different chromatic aspects of the original scene. This does not, however, require three times the channel capacity, as the resolution of the eye depends markedly on color. For example, in a red/green/blue (RGB) system, the green image might be transmitted at full resolution, the red at half resolution (one quarter on an area basis) and the blue at one fourth resolution (1/16 on an area basis.) Thus the color image would require only about a 30% increase in bandwidth over monochrome. Still better results would be obtained by transforming from RGB form into a different set of three components -- luminance (the intensity aspect of the stimulus, independent of color) and two chrominance signals. Both chrominance components could be reduced in resolution significantly, so that the incremental bandwidth required for color would be quite small. [USE FIGURE HERE]

Our present NTSC system, named after an industry group that proposed it, the National Television System Committee, was established as the US standard in 1953 primarily to achieve compatibility with the then-existing monochrome system. In this very clever system, luminance and chrominance information is transmitted as a single composite signal rather than as three separate components as in the earlier noncompatible field-sequential system. A band-sharing technique is used in which the two chrominance signals are mixed with the luminance signal, requiring them to be separated in the receiver and then converted back into RGB form for display on the picture tube. The luminance signal is much like a standard monochrome signal and can be received on unmodified monochrome receivers. Color receivers can also receive monochrome broadcasts properly.

The NTSC system and its European opposite numbers, PAL and SECAM, have spawned large and profitable industries. However, they are far from perfect. The horizontal color resolution is much lower than the vertical, resulting in horizontal smears of small, brightly colored areas. The luminance and chrominance signals interfere with each other. The color information not completely separated from luminance produces crawling serrations along the edges of large, brightly colored areas. Luminance components with the same frequency as color components can (and do) appear as false color patterns, especially in areas of high detail content. As a practical matter, the introduction of the color subcarrier

effectively reduced the luminance bandwidth, and thus the horizontal sharpness of the final image.

Most special effects require operations on signals in component form, necessitating repeated conversion between that form and the NTSC composite form. This exacerbates the problems due to chrominance/luminance band-sharing with consequent loss of quality. As TV equipment has gotten better, these defects have become more noticeable, leading recently to the introduction of component equipment, including cameras, VTR's and special effects generators, for studio use. Channel 7 in Boston, for example, does its electronic news-gathering production work in component format, translating to the NTSC composite format just once, after playback from a component-style VTR.

Compatible Improvements to the NTSC System

At the receiver. In view of the ample evidence that the receiver of the present system does not achieve the image quality to which its bandwidth consumption entitles it, engineers have expended some effort to make improvements without changing the nature of the transmitted signal. More sophisticated separation of the luminance and chrominance signals using special filters eliminates much of the problem, at least for stationary subjects, and regains some of the lost luminance resolution. The more interesting improvements come from the use of a frame store, a semiconductor memory component that holds one entire TV image. This permits the interpolation of additional scan lines between those transmitted. The simplest such schemes double the number of scan lines per sec, displaying 525 progressively scanned lines in each field of 1/60 sec. This reduces visibility of scan lines and virtually eliminates interline flicker from the displayed image. For stationary images, the interpolation is best done temporally, i.e., by using corresponding lines in successive frames, so as to preserve the vertical resolution, but this blurs moving objects. Some adaptive systems interpolate temporally only in stationary areas. In moving areas, they interpolate vertically, i.e., by using adjacent lines in the same field. [DIAGRAM]

The divorce of scanning standards of picture tubes from those of the camera and channel can be carried further. It is possible to go up to 1050 lines, either 30 fps interlaced or even 60 fps progressively scanned. This greatly reduces the visibility of the line structure and makes possible the use of quality enhancement techniques mentioned previously. Generally, these schemes average out the noise in areas that lack detail, where noise is most visible, and sharpen the image in detailed areas, where the inevitable increase in noise is not easily perceived. Many computer graphics displays are now operating at about these rates, producing images that seem to be printed on paper and pasted on the front of the picture tube. Such systems, of course, raise the cost of the receiver, but by less than one might expect. The frame memory is the key component. It can be made from 7 or 8 special, high-speed 256K memory

chips, at present less than $100. [CHECK CURRENT PRICES] Predictions of a $10 frame memory within a few years are common. It is also necessary to improve some of the analog circuitry, but this must be done for any high-definition receiver and will not be very costly in mass production.

Improvements at the transmitter. Just as separating the receiver's scanning parameters from those o the transmitted signal has solved some problems, another group of improvements depends on separating the camera's parameters from those of the channel. By using progressive scan in the camera at 60 fps and with at least 525, but preferably 1050 lines, it would be possible to derive a standard signal for transmission that is free from most of the present defects associated with interlace. Technically, these defects are called *aliasing,* or the masquerading of one spectral component for another, representing a different spatial pattern from that actually scanned. Aliasing can be reduced by filtering, which requires more image samples than are to be transmitted. Noise can also be reduced and sharpness enhanced by operating on the high-rate signal, but a limit is placed on the vertical sharpness because it might produce even worse interlace problems on receivers not equipped with frame memories.

In another class of improved systems, exemplified by CBS's DBS proposal, now shelved for economic reasons, the transmission consists of two signals: one can be readily received on existing receivers; the other is an enhancement signal that is added to the first signal to produce high-definition pictures on special receivers. In the CBS system, the camera operates at 1050 lines. Filtering followed by subsampling produces a 525-line picture for the compatible channel. A signal that is essentially the difference between the original and compatible signal is sent on the second channel. In this case, extra information for the sides of the images is also transmitted, so that the final image has a wider aspect ratio. A relatively inexpensive converter is needed in any case to deal with the C-band (?) satellite signal, and this unit would also effect the conversion to NTSC format. This system produces images of quality comparable to those of the NHK system, but has the added feature of near compatibility.

Beyond NTSC

Further enhancement of quality requires either adding more bandwidth or restructuring the video signal in some noncompatible manner. Several proposals for this kind of improvement involve separate transmission of color components, rather than using the NTSC composite band-sharing technique; they are collectively called multiplexed analog component (MAC) systems. Luminance/chrominance interference is eliminated by transmitting the components in separate frequency bands, or, preferably, in sequence, time-compressed, on each scan line. Compression and decompression are done digitally at both tranmsitter and receiver, but require only a few line memories, and not full frame memories. (Of course,

frame memories can also be used for additional benefit, if desired, just as in the improved NTSC systems mentioned.) [DIAGRAM] The image quality of typical MAC systems roughly equals that provided by the CCIR digital studio standard, which is very much better than normally achieved by PAL. The "MAC packet" systems espoused by the European Broadcasters' Union (EBU), also make provision for the transmission of multichannel digital audio plus a good deal of auxiliary data. It is anticipated that such signals would be transmitted by DBS. PAL-compatible signals, though of lower quality, could be derived for conventional earth-bound retransmission.

Much Better Systems for the Future

To have much better image quality — 35mm theater quality or better — with little or no bandwidth expansion beyond NTSC, we must use the channel capacity more efficiently. Such radically improved systems must, of course, also use much better cameras and displays, and may use some of the techniques already demonstrated for eliminating some of the defects of interlace. Any new system almost certainly would abandon the band-sharing scheme for chrominance transmission.

Better cameras and displays. It costs nothing in channel capacity to use improved cameras and displays. In interlaced systems, very high vertical resolution is not useful in these components since it worsens interline flicker. This struck me during a visit to Sony in 1981, where I first saw high-definition picture tubes designed for the NHK-1125 line system. In my innocence, I suggested that NTSC pictures must look really good on such tubes, but was told that, in fact, they looked terrible! In some cases, camera tubes would give higher definition merely if they were operated at a higher line number. In other cases, adaptive enhancement schemes of the kind already used in graphic arts could give better signals. Finally, progress in solid-state cameras will almost certainly eventually provide all the resolution desired. Some invention may be needed to improve or supplant picture tubes, as it is becoming more and more difficult to improve the resolution of the current designs further. In addition, the larger screens that are desirable will have to be made flatter to fit into today's living rooms. Work on small projection systems and on solid-state displays may solve the problem.

Separation of camera and display parameters from the channel. In addition to improving their basic resolution, the most important change for the camera and display is to operate them at substantially higher line and frame rates than the channel. The advantages of this approach are the elimination o visible line structure and of its temporal counterpart, flicker, the facilitation of optimum enhancement (by filtering) before converting to the channel signal, and the simplification of postproduction manipulation, including scan conversion. If carried far enough, these elements would become "transparent," so

that the quality of the final image would depend only on the information sent through the channel. Optimum sampling patterns for deriving the channel signal from the camera output must be considered as well as the filtering (interpolation) used to derive the signal for the display from the transmitted information.

Motion compensation. It appears possible to get better motion rendition with fewer transmitted frames per second. If the necessary receiver processing can be made practical, this is likely to be the most important avenue to the goal of better quality with little or no bandwidth expansion. The possibility depends on the fact that successive frames of a TV sequence *must* be very similar to give a satisfactory impression of continuous motion. [SEE BOX] If the frame-to-frame motion of each point in the image is known, then intermediate images can be calculated quite accurately from those frames that are transmitted. This is already done in computer graphics and in computer-generated cartoons, where it is called "in-betweening." In these cases, the motion is known *a priori*. In natural scenes, the motion must be calculated from the video information. Noise presents an obstacle to doing this with perfect accuracy. In addition, the reconstruction of the signal for the display from the sparse samples that would be sent through the channel is a formidable computational task that must be performed in each receiver. Difficult problems must be solved to make this technique practical, but the rewards for doing so are correspondingly large.

****

From these considerations, it is possible to discern the shape of a possible TV system of the future. The efficiency of bandwidth utilization, i.e., the image quality obtained in relationship to the channel capacity used, will be significantly improved compared with present systems. It will use excellent cameras and displays operated at very high line and frame rates — well over 1000 lines and perhaps as high as 100 fps — and probably progressively scanned. The camera signal will be appropriately filtered and sampled, probably in an offset pattern, giving higher spatial and lower temporal resolution than at present for transmission through the channel. These samples will be adaptively interpolated, using motion information either derived from the samples or transmitted directly (or both), up to a very high line and frame rate for display. There are additional possibilities for achieving greater efficiency including differential coding (only applicable to digital transmission) and multichannel transmission, in which the signal is divided into several channels of different types of spatial and temporal detail, each channel specially tailored to take advantage of specific perceptual characteristics. These latter methods require more research to prove their validity.

Worldwide Standards?

The present multiplicity of transmission standards — NTSC, PAL, SECAM, and their many variants — necessitate conversion in order to use, in one system, programs made in another. The conversion equipment is quite expensive at present and causes some loss of quality, especially for moving images and in cases where the frame rate must be changed. It should be noted that, in spite of this, the entertainment value of converted programs is undiminished. There is no evidence that technical production standards have any influence whatever on the salability of programs. However, if all countries wanted easy international transmission (it is not clear that they do) it would be advantageous to use a common standard.

It is equally important to be able to convert easily between film and video. Much video programming has always been derived from film productions, and many theatrical presentations, especially outside the US, are now originally produced for television. The most difficult conversion problem is the frame rate — 24 for film, and 25 or 30 fps for video in the 50- and 60-Hz countries. It is not possible to make theatre-quality films from NTSC or PAL because their spatial resolution is too low. Thus all productions intended for both the cinema and TV must now be produced on film, which has become the *de facto* international medium for exchange of programs.

With the advent of HDTV, it has become technically possible to use a new medium, namely high-definition TV, for both film and television production. This would be quite advantageous to producers, since video production is faster and cheaper than film, and editing is easier. Primarily for this reason, several groups have proposed the establishment of a worldwide TV production standard, naturally based on the NHK system, as it is the only system now available. It is not obvious that an HDTV standard for this essentially commercial purpose must be the subject of intergovernmental agreement, especially in the current deregulatory climate. There is much talk about avoiding another situation such as the existence of two competing formats, VHS and Beta, for home video cassettes. In the past most such problems were handled on a voluntary, nongovernmental basis. For example, The Society of Motion Picture and Television Engineers (SMPTE) has successfully coordinated standards for motion picture film. Proponents have also advanced a second argument, which has a more valid claim to the attention o governments. If a large proportion of TV-only producers, worldwide, were to adopt a common standard for production, it would also facilitate interchange of television programs, in the sense that they would not have to be converted at the production level.

Almost any 60-Hz interlaced standard would serve for this second objective, since most technical difficulties are associated with frame-rate conversion. Actually, the main reason why many industry leaders